社団法人　情報処理学会　研究報告
IPSJ SIG Technical Report

2003－ＮＬ－158　(19)
2003／11／7

多項分布と一様分布の混合分布による語義の事前分布の推定

新納　浩幸†　　佐々木　稔††

† 茨城大学工学部システム工学科　〒316–8511 茨城県日立市中成沢町 4-12-1
†† 茨城大学工学部情報工学科　〒316–8511 茨城県日立市中成沢町 4-12-1
E-mail: †shinnou@dse.ibaraki.ac.jp, ††sasaki@cis.ibaraki.ac.jp

あらまし　本論文は，多義語の曖昧性解消問題を対象に，語義の事前分布を多項分布と一様分布の混合分布により推定することを提案する．訓練データを利用する立場からは，多項分布であるか一様分布であるかを情報量規準を用いて選択できる．そして訓練データの事例数が多い場合には，多項分布が現実的であることを示す．また訓練データは利用できないとする立場からは，一様分布を仮定することが行なわれる．本論文の提案はこれらの融合案である．実験では Senseval2 の日本語辞書タスクを用いた．学習手法としては Naive Bayes 法と決定リストを用いた．各々の手法について，事前分布として多項分布，一様分布，およびそれらの混合分布を利用した場合の各々の正解率を調べた．実験結果から本手法の有効性を示した．

キーワード　事前分布，混合分布，多義語の曖昧性解消，Naive Bayes, 決定リスト

Estimation of prior distribution of word senses by mixture distribution of multinomial and uniform distributions

Hiroyuki SHINNOU† and Minoru SASAKI††

† Department of Systems Engineering, Ibaraki University　Nakanarusawa 4–12–1, Hitachi-shi, Ibaraki, 316–8511 Japan
†† Department of Computer and Information Sciences, Ibaraki University　Nakanarusawa 4–12–1, Hitachi-shi, Ibaraki, 316–8511 Japan
E-mail: †shinnou@dse.ibaraki.ac.jp, ††sasaki@cis.ibaraki.ac.jp

Abstract　In this paper, we estimate prior distribution by mixture distribution of multinomial and uniform distributions, for word sense disambiguation (WSD) tasks. In the case that we take advantage of training data, we can judge which is fitter multinomial distribution or uniform distribution, by using Akaike Information Criterion (AIC). As the result, we show that multinomial distribution is selected if training data is large. On the other hand, there is another view that we should not use training data to estimate prior distribution. We explain this view. These two views are harmonized by our mixture model. In experiments, we used the Japanese dictionary task of Senseval2. As learning methods, we used Naive Bayes and decision list. For each learning method, we evaluated precisions of the task in each case that using multinomial distribution, uniform distributions and mixture distribution as prior distribution. As the result, we showed our method is effective.

Key words　prior distribution, mixture distribution, word sense disambiguation, Naive Bayes, decision list

1. まえがき

ある種の帰納学習手法を利用して多義語の曖昧性解消を行う場合，語義の事前分布を推定する必要がある．語義の事前分布として通常は多項分布を用いるが，本論文では多項分布と一様分布の混合分布を用いることを提案する．

多義語の曖昧性解消は分類問題の一種であり，様々な機械学習手法を用いて解決できる．ここでは Naive Bayes 法と決定リストを取り上げる．両手法には共通して語義の事前分布の推定が必要となる．

基本的に確率統計的な手法では，事例 x がクラス c である確率 $P(c|x)$ を求めることで分類問題を解決する．$P(c|x)$ はベイズの定理から，以下のように変形できる．

$$P(c|x) = \frac{P(c)P(x|c)}{P(x)}$$

この $P(c)$ が語義の事前分布に対応する．Naive Bayes 法 [2] でも上式の右辺を求めるアプローチをとるために，語義の事前分布が必要になる．また，決定リスト [3] はクラスを識別するための素性をその予測力の順に並べたものである．実際の識別では上位の素性から順に，その素性が文脈に存在するかどうかを調べ，もしその素性が存在すれば，その素性に対応したクラスを識別結果し，存在しなければ次の素性を調べる．そして，リスト中のすべての素性が文脈に存在しない場合，default 規則が適用される．default 規則は一般に最大の出現確率をもつクラスを識別結果とする．このため決定リストにおいても語義の事前分布が必要になる．

語義の事前分布は訓練データの語義の相対頻度から求めることが通常行なわれている．これはデータが多項分布から発生していることを仮定し，パラメータを最尤推定していることに対応する．しかしデータが多項分布であるという仮定は必ずしも正しいとは限らない．本論文では訓練データがある程度大きければ，多項分布を仮定することは妥当であることを情報量規準（Akaike Information Criterion 以下 AIC）[5] を用いて示す．一方，事前分布は訓練データからは求められないとする考え方も存在する．この考え方に従うと事前分布は一様分布として設定するほかない．本論文では多項分布の考え方と一様分布の考え方を融合する形として，両者の混合分布を事前分布とすることを提案する．

実験では Senseval2 の辞書タスク [8] のデータを用い，多項分布，一様分布およびそれらの混合分布（提案手法）で推定した場合の正解率を測り，提案手法の有効性を示す．

2. 事前分布の必要性

2.1 Naive Bayes 法の場合

今，ある多義語 w の語義の集合を $C = \{c_1, c_2, \cdots, c_m\}$ とおき，ある文脈 x 上で現れた w の語義を識別することを考える．確率統計的には $P(c_i|x)$ を最大にするような語義 c_i を識別結果とすればよい．ベイズ推定ではベイズの定理を用いることで，

$$\begin{aligned}\arg\max P(c_i|x) &= \arg\max \frac{P(c_i)P(x|c_i)}{P(x)} \\ &= \arg\max P(c_i)P(x|c_i)\end{aligned}$$

の変形を行い，$P(c_i)P(x|c_i)$ を最大にするような c_i を求める問題に変形する．ここで $P(c_i)$ の部分を事前分布，$P(x|c_i)$ の部分を事後分布と呼ぶ．

事前分布 $P(c_i)$ の推定は後述することにし，ここでは事後分布 $P(x|c_i)$ の推定方法を説明する．多語義 w が現れた文脈 x は素性リスト $x = (x_1, x_2, \cdots, x_n)$ として表せる．通常，与えられている訓練データのみからは $P(x|c_i)$ の推定は困難である．ここで各素性は独立と考えると以下の変形が可能になる．

$$P(x|c_i) = \prod_{j=1}^{n} P(x_j|c_i)$$

$P(x_j|c_i)$ の推定は容易であることから，結果として $P(x|c_i)$ が計算できる．この「各素性は独立」という仮定に基づき，上記の式の変形から $P(x|c_i)$ を計算するアプローチを Naive Bayes 法という [2].

以上より，Naive Bayes 法は事前分布と事後分布の積により識別を行うので，事前分布の推定が必要になることは明らかである．

2.2 決定リストの場合

決定リストは訓練データをもとに以下の手順で作成される表である．

step 1　訓練データから素性とクラスの組の頻度を調べる．

訓練データはクラスの付与された事例の集合である．事例 x は素性リスト $x = (x_1, x_2, \cdots, x_n)$ として表せる．この事例 x のクラスが c であるとき，この事例から x_i と c の組 (x_i, c) の頻度に 1 を足す．この操作をを訓練データ中の全事例の全素性について行う．

step 2　素性の予測力と識別クラスを導く．

(x_i, c) の頻度が f_c であった場合，f_c の最大値を与える $c = \hat{c}$ が素性 x_i に対する識別結果となる．またそのときの予測力 $est(x_i)$ も求める．$est(x_i)$ の定義については後述する．

また $default$ という特別な素性も設定する．これは訓練データ中で，クラス c の頻度が S_c であった場合，S_c の最大値を与える $c = c_{def}$ が $default$ に対する識別結果である．$default$ の予測力 $est(default)$ の定義も後述する．

step 3　予測力の順に並べる

全ての素性と識別結果の組を予測力の大きい順に並べる．これによって作成できた表が決定リストである．ただし素性 $default$ の予測力よりも小さなものは表から外す．

実際の識別はあるテスト事例が与えられた時に，決定リストの上位の素性から順にその素性がテスト事例 (素性リスト) 中に存在するかどうかを調べる．もし存在すれば決定リストのその素性に対する識別結果が出力となる．もし存在しなければ次の素性に移る．もしも決定リスト中のすべての素性がテスト事例中に存在しない場合には，$default$ に対する識別結果が出力となる．

step 2 の予測力の定義は，一般に以下の対数尤度比が用いられる．

$$est_1(x_i) = \log \frac{P(\hat{c}|x_i)}{1 - P(\hat{c}|x_i)}$$

$P(\hat{c}|x_i)$ は step 1 で得られた頻度の情報から計算できる．また $default$ の予測力 $est(default)$ は以下で定義される．

$$est_1(default) = \log \frac{P(c_{def})}{1 - P(c_{def})}$$

予測力の定義は対数尤度比を用いずに，以下のように，直接，確率の値を与える場合もある．

$$est_2(x_i) = P(\hat{c}|x_i)$$

$$est_2(default) = P(c_{def})$$

上記の対数尤度比と確率は 1 対 1 に対応し、大小関係が保たれる．

$$est_1(a) \geqq est_1(b) \Leftrightarrow est_2(a) \geqq est_2(b)$$

そのためどちらの定義を用いても決定リストに差は生じず(注1)，識別の結果に影響を与えない．

ここで注目すべきは $est(default) = P(c_{def})$ である．つまり，決定リストの場合，$default$ の予測力は語義の事前分布を利用している．

3. 事前分布の推定

3.1 訓練データを利用する立場

事前分布は訓練データ中のクラスの相対頻度から求めることが一般に行われている．これは事前分布として多項分布を仮定し，そのパラメータを最尤推定していることに相当する．しかし多項分布を仮定することが妥当であるかどうかは確認が必要である．

カテゴリカルなクラスに対する分布を仮定する場合，多項分布あるいは一様分布を用いるのが自然である．どちらの仮定がより適切かはモデル選択の理論から判断できる．ここではAIC [5] を用いて多項分布あるいは一様分布のどちらの分布を仮定するのがよいのかを判断する．

3.1.1 AIC

AIC とはモデル選択の理論である．データが与えられたときにそのデータを発生しているモデルとして M_1 と M_2 が想定できる時に，そのどちらのモデルがより適切かを判断する．

AIC では，概略，モデルの最大対数尤度とそのモデルのパラメータ数の差が近似的に期待平均対数尤度の不偏推定量になることに着目して，モデル M のモデル選択規準 $AIC(M)$ を以下のように定義する．

$$AIC(M) = -2((M \text{ の最大対数尤度}) - (M \text{ のパラメータ数}))$$

モデルとして M_1 と M_2 が想定できるとき，上式より $AIC(M_1)$ と $AIC(M_2)$ を計算し，その値が小さい方を適切なモデルとして選択する．

3.1.2 AIC による多項分布と一様分布の比較

クラスの集合を $C = \{c_1, c_2, \cdots, c_m\}$ として $P(C = c_i)$ の分布が多項分布であると仮定し，$P(C = c_i) = p_i$ とおく．またこのモデルを M_1 とおく．一方，$P(C = c_i)$ の分布は一様分布であると仮定したモデルを M_2 とおく．

今，訓練データは n 事例あるとしてクラス c_i の事例が n_i 個あるとする．このとき，M_1 の最大対数尤度は以下で与えられる．

$$\sum_{i=1}^{m} n_i \log n_i - n \log n$$

また M_1 のパラメータ数は $m-1$ であるので，

$$AIC(M_1) = -2(\sum_{i=1}^{m} n_i \log n_i - n \log n - m + 1)$$

となる．一方，M_2 の最大対数尤度は以下で与えられる．

$$\sum_{i=1}^{m} n_i \log \frac{1}{m}$$

また M_2 にパラメータはないので，

$$AIC(M_2) = -2 \sum_{i=1}^{m} n_i \log \frac{1}{m}$$

となる．

今，簡単なケースとして $m = 2$ の場合，つまり語義が 2 つの場合を考えてみる．表 1 は n に対して，AIC によりモデル M_2 つまり一様分布の方が選択される n_1 の範囲とその割合を示している．例えば $n = 100$ では，$n_1 = 43 \sim 57$ のときに M_2 が妥当だと判定される．n_1 は 0 から 100 までの 101 種類の値を取り得ることができ，43 から 57 の 14 種類は割合的には 14.85% (= 14/101) である．この表から明らかなように n が大きくなればなるほど，M_2 が選択されることはなくなる．

表 1 一様分布が選択される範囲と割合

事例数	範囲	割合
50	21 ～ 29	17.65%
80	34 ～ 46	16.05%
100	43 ～ 57	14.85%
150	67 ～ 83	11.26%
200	91 ～ 109	9.45%
300	138 ～ 162	8.31%

実際の語義識別では，訓練データの事例数 n は 100 以上はあるだろうし，$m \geqq 3$ の場合も多い．そのような場合にはよりいっそう，$M_1 < M_2$ の傾向がある．そのため訓練データを利用して事前分布を推定するには，AIC を用いてきっちりと判定してもよいが，大雑把に，多項分布を用いてもよいと考えられる．

3.2 訓練データを利用しない立場

本来，訓練データから事前分布を推定することができると考えるのは，訓練データが全体のデータからのランダムサンプルであることを仮定している．例えば，識別が難しいような事例を集中的に訓練データとして収集した場合には，事前分布を訓練データから推定することに意味はない．このように明らかに作為的な操作を経なくても，新聞データから事例を集めれば，新聞記事特有の文書ばかりであり，そこにはやはりランダム性が損なわれている．また，本質的ではないが，Senseval のようなコンテストで，テストデータに対する正解率をあげることだけを考えてみる．この場合，仮に訓練データがランダムサンプルであったとしても，テストデータがランダムサンプルでなければ，訓練データから事前分布を推定することは，テストデータに対する識別精度の向上につながらない．

このように疑った態度を取ると，訓練データから事前分布を推定することは不可能であり，結局，一様分布として取り扱う以外に方法はないことがわかる．

(注1)：実際はスムージングを行なうのでその方法によって若干の差が生じることもある．

3.3 混合分布による推定

前節からある程度の訓練データが用意され，しかも，訓練データを活用する態度をとれば，事前分布は多項分布としてモデル化できる．また訓練データを疑う態度をとれば，事前分布は一様分布としてモデル化する以外ない．

そこで本論文では，これらの考えを調和する形で，語義の事前分布として多項分布と一様分布との混合分布を用いることを提案する．

$P_1(c)$ を多項分布，$P_2(c)$ を一様分布としたとき，多項分布と一様分布の混合分布は

$$P_3(x) = \alpha P_1(c) + (1-\alpha)P_2(c)$$

と表せる．ここで α は 0 以上 1 以下のある定数であり，事前分布を推定する際に訓練データを活用する度合に対応する．本論文の提案では $\alpha = 0.5$ となる．

実際の推定は，訓練事例の数を n，訓練事例の中でクラス c の事例の数を n_c，クラスの種類数を m としたとき，

$$P_1(c) = \frac{n_c}{n}$$

$$P_2(c) = \frac{1}{m}$$

なので，

$$P_3(c) = 0.5(\frac{n_c}{n} + \frac{1}{m})$$
$$= \frac{n_c m + n}{2mn}$$

となる．

4. 実　　験

4.1 日本語辞書タスク

Senseval2 の日本語辞書タスクを用いて本手法の有効性を確認する．

日本語辞書タスクは一般的な多義語の曖昧解消タスクである．名詞 50 単語と動詞 50 単語の多義語が用意され，訓練データは 1 単語平均して名詞 177.4 事例，動詞 172.7 事例用意されている．各単語に対するテストデータとして 100 事例が提供されている．

評価は事前分布として，(a) 多項分布，(b) 一様分布，(c) それらの混合分布 (提案手法) を用いた場合に，Naive Bayes 法および決定リストにより学習できた規則のテストデータに対する識別の正解率により行なう．また正解率の測り方は部分点も与える mixed-grained scoring [8] 方式を使う．

4.2 素性の設定

学習の際に利用した属性は以下の 6 種類である．

e1	直前の単語
e2	直後の単語
e3	前方の内容語 2 つまで
e4	後方の内容語 2 つまで
e5	e3 の分類語彙表の番号
e6	e5 の分類語彙表の番号

例えば，語義識別対象の単語を「記録」として，以下の文を考える（形態素解析され各単語は原型に戻されているとする）．

過去/最高/を/記録/する/た/。

この場合,「記録」の直前，直後の単語は「を」と「する」なので，‘e1=を’，‘e2=する’となる．次に,「記録」の前方の内容語は「過去」,「最高」なので，ここから「記録」に近い順に 2 つ取り，‘e3=過去’，‘e3=最高’が作られる．またここでは句読点も内容語に設定しているので,「記録」の後方の内容語は「する」と「。」となり，‘e4=する’，‘e4=。’が作られる．次に「最高」の分類語彙表 [4] の番号を調べると，3.1920_4 である．ここでは分類語彙表の 4 桁目と 5 桁目までの数値をとることにした．つまり‘e3=最高’に対しては，‘e5=3192’と‘e5=31920’が作られる．同様に「過去」の分類語彙表の番号 1.1642_1 から‘e5=1164’と‘e5=11642’が作られる．次は「する」の分類語彙表を調べるはずだが，ここでは平仮名だけで構成される単語の場合，分類語彙表の番号を調べないことにした．これは平仮名だけで構成される単語は多義性が高く，無意味な素性が増えるので，その問題を避けたためである．もしも分類語彙表上で多義になっていた場合には，それぞれの番号に対して並列にすべての素性を作成する．

結果として，上記の例文に対しては以下の 10 個の素性が得られる．

e1=を，e2=する，e3=最高，e3=過去，
e4=する，e4=。，e5=3192，e5=31920，
e5=1164，e5=11642

上記の例文をデータ x としておくと，データ x はこの 10 個の素性を要素として持つリストとして表せる．

x = (e1=を, e2=する, e3=最高, e3=過去,
e4=する, e4=。, e5=3192, e5=31920,
e5=1164, e5=11642)

4.3 実験結果

まず Naive Bayes 法による結果を示す．事前分布として，(a) 多項分布，(b) 一様分布，(c) それらの混合分布 (提案手法) のそれぞれを利用した場合の正解率を以下に示す．

表 2　Naive Bayes 法

	多項分布	一様分布	混合分布 (提案手法)
名詞	0.7692	0.7692	0.7734
動詞	0.7804	0.7835	0.7814
合計	0.7748	0.7763	0.7774

次に決定リストによる結果を示す．事前分布として，(a) 多項分布，(b) 一様分布，(c) それらの混合分布 (提案手法) のそれぞれを利用した場合の正解率を以下に示す．ただし事前分布として一様分布を採用した場合，default の識別クラスが確定できない．このため事前分布として一様分布を採用した場合，default の識別クラスは頻度が最大のクラスを選択することにした．

表 3　決定リスト

	多項分布	一様分布	混合分布 (提案手法)
名詞	0.7618	0.7642	0.7640
動詞	0.7761	0.7777	0.7781
合計	0.7690	0.7710	0.7711

Naive Bayes 法も決定リストも，事前分布として混合分布を用いたものがわずかではあるが，最も精度が高く，本手法の有効性が示された．

5. 考　　察

5.1 多項分布への決めうち

訓練データから一様分布か多項分布かを判断する際には AIC を用いればよい．ただしここでは多項分布に決めうちしている．実際に，辞書タスクの各単語で AIC を用いて一様分布が選ばれるか，多項分布が選ばれるかを調べた．結果は，名詞「市民」だけが一様分布と判定され，その他はすべて多項分布と判定された．

実際の「市民」は語義が 2 つある．(a)『市の住民』と (b)『国家への義務、政治的権利を有する国民。公民。』である(注2)．訓練データ中の (a) の語義の頻度は 56，(b) の語義の頻度は 51 であった．多項分布とした場合，$P(c_a) = 0.5234$ であり，一様分布 $P(c_a) = 0.5000$ とした場合とほとんど差がない．実際に一様分布でも多項分布でも「市民」に関する正解率に変化はなかった．

訓練データの事例数が多い場合は，AIC を用いるとほとんどの場合多項分布と判定される．仮に一様分布と判定されるにしても，多項分布との差はほとんどなく，正解率を大きく下げることはないと思われる．

(注2)：実際は 3 つ目の語義 (c)『近代史で、前代の貴族・僧侶に代わって政治的権力を得た階級。ブルジョア。』の語義もあるが，訓練データ中には現れていないため，ここではこの語義はないと考えた．一様分布を仮定する場合でも，このように訓練データ中に現れない語義は存在しないものとしている．

5.2 アンサンブル学習

アンサンブル学習とは複数の学習手法を組み合わせる学習手法であり，様々なタスクにおいて成果をあげている [1]．アンサンブル学習では複数の学習手法を準備し，対象の分類問題に対してそれぞれの学習手法を用いて分類器を学習する．テストデータに対しては，作られた複数の分類器の出した識別結果を総合して最終的な識別を行う．これは，個々の学習手法の弱い部分を補う形になるために，得られる識別精度は個々の学習手法単独の識別精度よりも高くなる [6]．

本論文で示した事前分布を多項分布と一様分布の混合分布に設定することは，アンサンブル学習とみなすことができる．

Naive Bayes で事前分布として多項分布 $P_1(c)$ と一様分布 $P_2(c)$ を考える．前述したように，分類問題は $P(c_i|x)$ を最大にする c_i を求めればよい．ベイズの定理から，事前分布を多項分布とした場合には，

$$\arg\max P_a(c_i|x) = \arg\max P_1(c_i)P(x|c_i)$$

であり，事前分布を一様分布とした場合には，

$$\arg\max P_b(c_i|x) = \arg\max P_2(c_i)P(x|c_i)$$

である．この 2 つの分類器を組み合わせる．確率が導かれる分類器では単純に確率の総和を取ることで組合せが可能なので，上記 2 つをアンサンブルした場合には以下となる．

$$\begin{aligned}\arg\max P(c_i|x) &= \arg\max\left(P_a(c_i|x) + P_b(c_i|x)\right)\\ &= \arg\max(P_1(c_i) + P_2(c_i))P(x|c_i)\\ &= \arg\max\frac{P_1(c_i) + P_2(c_i)}{2}P(x|c_i)\end{aligned}$$

$(P_1(c_i) + P_2(c_i))/2$ は多項分布と一様分布の混合分布を示しているので，本手法を用いた Naive Bayes 法は，事前分布を多項分布とした Naive Bayes 法と事前分布を一様分布とした Naive Bayes 法とのアンサンブル学習となっていることがわかる．

5.3 決定リストにおける defalut の予測力

決定リストにおいて事前分布は default 規則に対応する．本手法では default に対する識別結果に変更はないので，一見，本手法を用いてもテストデータに対する識別結果に影響はないようにも思える．しかし実際はそうではない．本手法の場合，default の予測力が若干低くなる．このため，多項分布のときには default の予測力よりも小さいためにリストから削除されていた素性がリストに登場してくる．この部分が正解率に若干の影響を与えてくる．

決定リストは default 規則近辺の素性の扱いが重要である．決定リストの上位に位置する素性はその予測力も大きく，上位部分の素性を使った識別で誤ることは少ない．問題は下位部分の素性である．下位部分の素性の識別精度をあげることは，難しい課題であると思われる．

1 つの方法として本手法のように deafult の予測力を低く見積もり，有用な素性をリストに含めることである．また関連し

て低頻度の素性の扱いも重要である．一般に決定リストでは，ある頻度以下の素性はリストに含めないと言う間引き処理を行なう．ここでの実験でも頻度 1 のものは間引いた．間引き処理を行なわず予測力をつけること [7] も考えられるが，その場合には本手法のように deafult の予測力が重要になると思われる．

6. おわりに

本論文では多義語の曖昧性解消問題を想定して，語義の事前分布を多項分布と一様分布の混合分布による推定することを提案した．訓練データを利用する立場からは AIC を用いて多項分布あるいは一様分布を選択できることを示し，訓練データの事例数が多い場合には多項分布が現実的であることを示した．また訓練データは利用できないとし，一様分布を仮定する考え方も紹介した．提案手法はこれらの融合案である．実験では Senseval2 の日本語辞書タスクを用い，事前分布として多項分布，一様分布，およびそれらの混合分布を用いた場合の各々の正解率を調べた．実験結果から本手法の有効性を示した．今後は決定リストの defalut 規則近辺の素性をどのように扱えばよいかを考えたい．

文　献

[1] Ethem Alpaydm. Techniques for combining multiple learners. In *Engineering of Intelligent Systems*, Vol. 2, pp. 6–12, 1998.

[2] Tom Mitchell. *Machine Learning*. McGraw-Hill Companies, 1997.

[3] David Yarowsky. Decision Lists for Lexical Ambiguity Resolution: Application to Accent Restoration in Spanish and French. In *32th Annual Meeting of the Association for Computational Linguistics*, pp. 88–95, 1994.

[4] 国立国語研究所. 分類語彙表. 秀英出版, 1994.

[5] 坂元慶行, 石黒真木夫, 北川源四郎. 情報量統計学. 共立出版, 1993.

[6] 上田修功, 中野良平. アンサンブル学習における汎化誤差解析. 電子情報通信学会論文誌, Vol. J80-DII, No. 9, pp. 2512–2521, 1997.

[7] 鶴岡慶雅, 近山隆. ベイズ統計の手法を利用した決定リストのルール信頼度推定法. 自然言語処理, Vol. 9, No. 3, pp. 3–19, 2002.

[8] 白井清昭. SENSEVAL-2 日本語辞書タスク. 自然言語処理, Vol. 10, No. 3, pp. 3–24, 2003.